MANAGEMENT INFORMATION SYSTEMS

Ethics in Business Data Mining

Greg Smith, Ph.D.
Mentor: David Burns, DBA (Marketing)

For the Spring 2007 semester, I was asked to teach an introductory course in data mining for undergraduate business majors (INFO329/MKTG329 Data Mining). This was the second time a data mining course was included in the curriculum, but the first time in three years. The challenge was designing a course that provided both practical applications and ethical considerations. The course introduced students, for the first time, to emerging techniques that are ethically neutral, but potentially harmful when humans are left to the design.

What Is Data Mining?

For purposes of an introductory data mining class, data mining is the employment of machine learning to analyze data from many different perspectives and the summarization of the relationships identified. For business purposes, this information is to be used to improve overall business performance.

We see the wonderful results of data mining everyday. But, as helpful as data mining can be in improving our lives (i.e. better shopping experiences, election coverage, Google, etc.), it can have the opposite effect if mishandled or used in a harmful way (i.e. spam, junk mail, pop-up ads, etc.).

Course Vision

My vision of this course was to present and discuss data mining technologies and their applications in an effort to better support business decisions. I allotted time at the end of the term for ethical considerations to the practice through readings, lecture, and discussion.

Upon completion of this course, my hope was that students were able to:

- Understand popular data mining techniques, how to apply them, and when they are applicable
- Utilize a state-of-the-art commercial data mining package
- Apply popular data mining techniques to solve "real-world" problems
- Recognize the ethical and social impacts of data mining on our society.

Original Course Ethics Component

The original course component covered one class period

Readings

The course reading on ethical data mining was drawn directly from our class text and briefly covered two aspects:

- The hazards of data mining, specifically governmental uses and misuses primarily since September 11, 2001,
- Web based data mining issues.

Lecture

The lecture component tied together ethical issues of data mining business applications. Several "real-world" examples of socially responsible and irresponsible data mining were presented along with the tools used to develop the new knowledge.

Discussion

The discussion focused on how to be socially aware and responsible data miners who are attuned to the effects of data mining on our daily lives.

Redesigned Course Ethics Component

While the original course ethics component was a vital piece, I found my approach insufficient. I thought it necessary to strengthen the component with additional student-centric content for the following year's section. So, to enable the change for the Spring 2008 semester, I surveyed an undergraduate section of business statistics, similar in student demographic characteristics to those in an undergraduate data mining class, about their views on data mining ethics from a shopper's perspective on easily recognizable topics. Through conversations with David Burns, I designed a short online questionnaire. The following is a sample of survey questions.

Shopper Cards

- Do you use a store shopper card for discounts, such as a grocery plus card?

    a) If so, did you read the user agreement?
    b) Are you aware of how your store transaction data is being used?
- Do you know if personal information (including: who you are, what you buy, etc.) is being "sold"?

Online

- Have you ever purchased anything from Amazon.com?
    a) If so, did you read the user agreement?
    b) Have you taken advantage of their personalized suggestions?
- Have you ever filled out a web form (online form) of any sort to gain access to information you needed (online newspapers, free downloads, websites such as Myspace, etc.)?
    a) If so, did you read the user agreement?
    b) Have you recommended other individuals to this site?

My intent was to draw basic feelings, knowledge, and current practices of students when using services designed for data mining to improve the ethics component.

As with the original course ethics component, I allotted time at the end of the term for redesigned component. The redesigned component stretched over a week of class and allowed for a natural wrap-up of the course.

Readings

In addition to content drawn directly from our class text, I was able to assign suitable readings for the component that included practical pieces tailored to student consumers.

Lecture

The revised lecture continued to tie together ethical issues of "real-world" data mining business applications. In addition, I invited guest speakers from dunnhumby, inc. to present on their company, its vision and practice, and the considerations that it has for consumers in the United States and around the world. The experience added depth to students' knowledge base and exposed them to current applications and social limitations of data mining.

Discussion

The revised discussion expanded the view of socially responsible data mining. In addition, we viewed the University of Notre Dame' Mendoza College of Business 2007/2008 Berges Lecture Series in Business Ethics discussion titled: "Data Mining: Business, Ethical and Societal Considerations: A Panel Discussion" from September 11,

2007 to the component. This first-of-its-kind forum covered data mining ethics from a management, marketing, and information systems perspective.

Ethics Assignment (New)
With a growing mental tool-box of data mining ethical knowledge, students were challenged to consider a current data mining opportunity for Xavier University as several off-campus businesses (Chipotle, Donatos, CVS, etc.) started accepting ALL CARDS for purchases. The students were asked to ponder the potential mining opportunities for Xavier University through transactional data that could be captured and how it could aid both the University and its students. The students were asked to address the situation as both socially responsible and irresponsible practitioners.

INSTRUCTOR REFLECTIONS
The change to the component presented a unique challenge as the field is still in its infancy with discussions and literature on ethics sparsely available. The pieces that I included broadened the in-class discussion from the original component and exposed the students to new considerations for knowledge discovery. The assignment gave students an opportunity to see data mining from both a practitioner and consumer standpoint. As a group they appeared to display a keen sense of social responsibility from both sides. I plan to use the redesigned component for upcoming classes and adapt the piece for the graduate curriculum.